Meaningful application of the new 454 large scale pyrosequencing technology (Roche GS-FLX 454) to the identification of microsatellites for small-scale research projects

M. Jeanneau[1], M-C Bon[1], and J-F Martin[2].

1. European Biological Control Laboratory USDA-ARS and 2. Centre de Biologie et de Gestion des Populations (CBGP) - SupAgro Campus International de Baillarguet, CS90013 Montferrier sur Lez, 34988 St. Gély du Fesc, France.

mjeanneau@ars-ebcl.org, mcbon@ars-ebcl.org, martinjf@supagro.inra.fr.

INTRODUCTION

Microsatellites (simple sequence repeats) are DNA sequences that consist of tandem repeats of 1-6 nucleotides (Figure 1). Because of high levels of polymorphism, ease to use and co-dominance, they are generally seen as the most pertinent markers to study at a fine scale level the genetic structure and demographic history of invasive weed populations. However, their development using the classical microsatellite-library by enrichment remain typically time consuming and labor intensive, impeding their generalization at the level of small-scale research projects. We considered in the present study the extent to which the new generation sequencing such 454 Life Sciences/Roche GS-FLX pyrosequencing based-technology, could lead to a rapid, more efficient and less costly way to identify microsatellites for a small-scale research project than the classical microsatellite-library by enrichment.

This new method was first developed within the framework of a multi-partner pilote project between the French Institut National de la Recherche Agronomique (INRA), AIP Bioressources EcoMicro), the R&D department of Genoscreen, Montpellier SupAgro and the University of Provence. To provide a broadly applicable test during the project, genomic DNA was sourced from Plant, Fungus, Arthropod, Nematod and Stramenopile kingdoms and numerous unrelated taxa within Arthropods including our target. Development of microsatellites has been streamlined by optimizing the numerous steps in microsatellite identification and subsequent sequencing throughput, to make the 454 process even more successful. The flowchart of both the classical microsatellite-library method and this improved 454 technology that has recently been published (Malausa et al., 2010) are presented here.

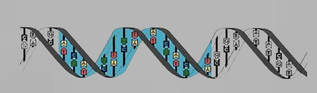


Figure 1 : Microsatellite (CA repetition).

DNA SOURCE:

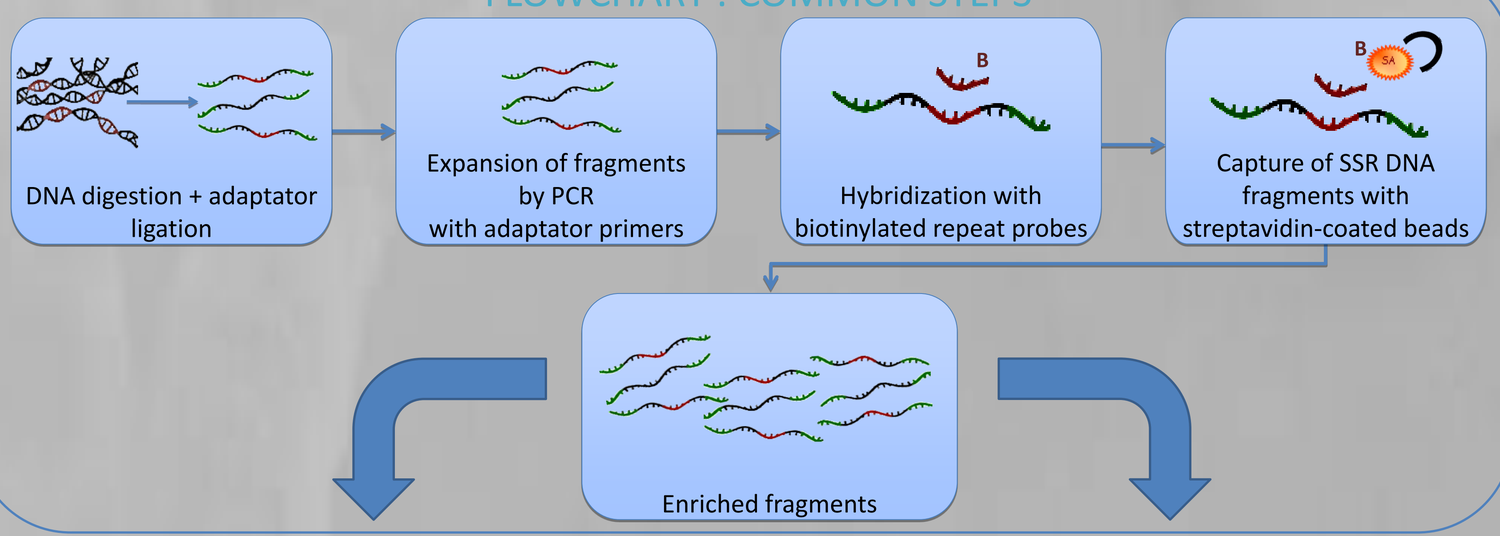
The DNA was sourced from a non model taxa i.e. a root gall weevil, *Ceutorhynchus assimilis* (Coleoptera: Curculionidae) that is one natural enemy of the invasive weed, *Lepidium draba* sp *draba* (Brassicaceae) (Figure 2). DNA was extracted using the Qiagen DNeasy Tissue DNA extraction kit.

Figure 2 : *Ceutorhynchus assimilis* © JEANNEAU M.

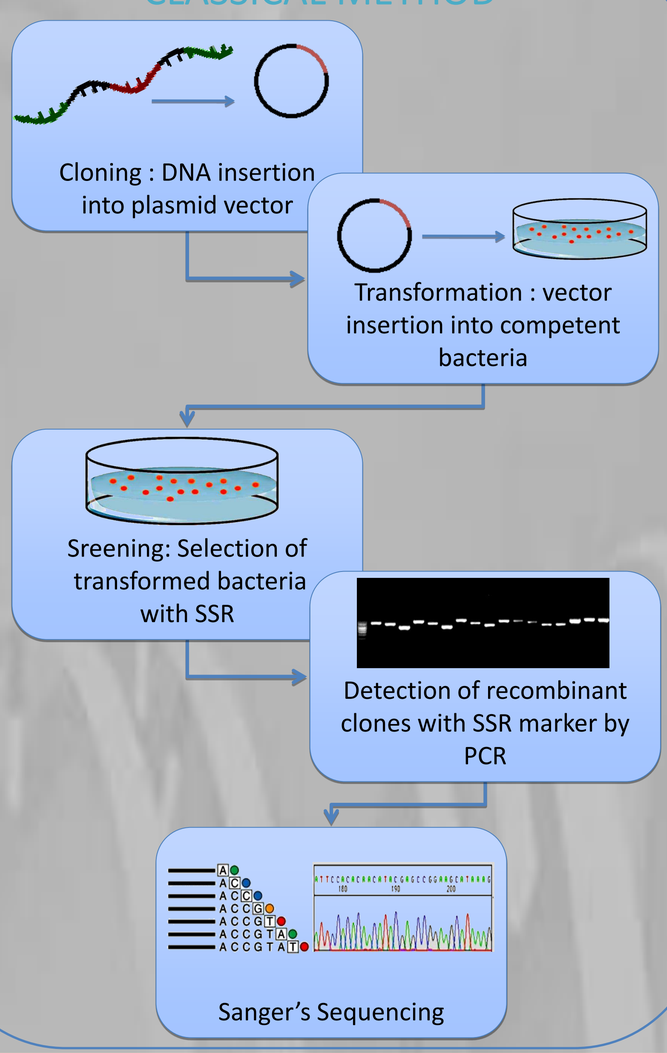
CLASSICAL METHOD: enriched library & Sanger sequencing

For construction of the enriched DNA library, we followed Kijas et al., (1994). 1,5 µg of genomic DNA was digested with *Rsa1* restriction enzyme. Digested fragments were ligated to *Rsa* adaptators and PCR amplified. PCR products were enriched with Streptavidin-coated magnetic beads (Promega) and 3'-biotinylated $(TC)_{10}$ or $(TG)_{10}$. Enrichments were PCR amplified and cloned. A total of 509 recombinant clones for the two libraries were picked randomly and tested by PCR for detection of microsatellites. Inserted fragments of 500 to 1,000bp in 96 clones were sequenced using the Sanger 's technology.

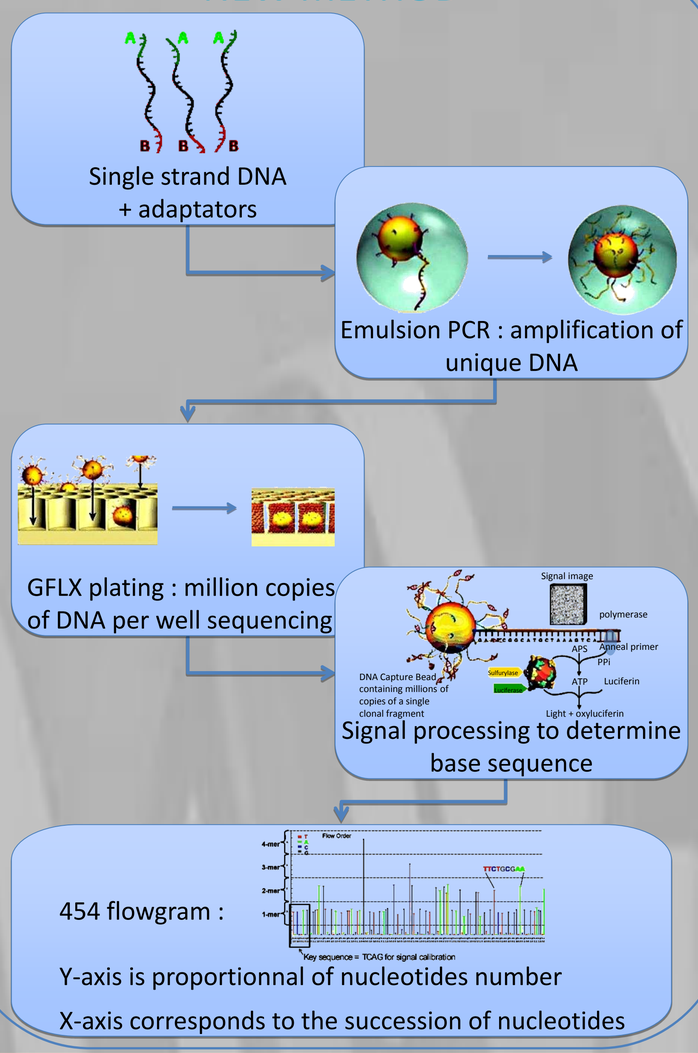
FLOWCHART : COMMON STEPS



CLASSICAL METHOD



NEW METHOD



NEW METHOD: 454-GS-FLX Titanium pyrosequencing of enriched library

1 µg of genomic DNA was digested by *Rsa1* restriction enzyme. Digested fragments were ligated to *Rsa* adaptators and PCR amplified. PCR products were enriched with Streptavidin-coated magnetic beads (Promega) and 3'-biotinylated $(AG)_{10}$, $(AC)_{10}$, $(AAC)_8$, $(AGG)_8$, $(ACG)_8$, $(AAG)_8$, $(ACAT)_6$, $(ATCT)_6$. Subsequent enrichments were PCR amplified.

The next steps were specific to the 454 GS-FLX technology and carried out as described by the manufacturer (Roche Applied Science). This implies fragment end polishing, adaptator ligation (A and B), addition of Multiplex identifiers (MIDs), library immobilization, fill-in reaction and isolation of a library of a single strand DNA + adaptors. During an emulsion PCR, each fragment is then mixed with beads carrying on their surfaces oligos complementary to the adaptators. With a PCR amplification occurring within each droplet, each bead is bound with millions of copies of a unique DNA (equivalent to clonal amplification). SstDNA library after addition of DNA polymerase and a cocktail of sulfurylase and luciferase was sequenced by the Genome sequencer FLX.

Using extension reaction by DNA polymerase just same as normal PCR, bases will be loaded in a fixed order of T -> A -> G -> C. Sulfurylase generates ATP using pyrophoric acid, which is made in polymerase extension reaction, as a substrate. Luciferase reacts to generate a light signal using this ATP and Luciferin as substrates. This chemiluminescent signal is recorded by the CCD camera.

DATA ANALYSIS

Sequences obtained in both protocols were analyzed using the bioinformatic QDD pipeline (Meglécz et al., 2010). Sequences shorter than 80bp and containing microsatellite motifs shorter than four repeats for any motif of 2 to 6bp were discarded. Sequences with significant Blasts hits but with flanking region identity levels below 90% were discarded. Primers were designed automatically by Primer3 within QDD.

CRITERIA	CLASSICAL METHOD	NEW METHOD
DNA quantity	> 1,5 µg	> 2 µg
Number of raw sequences	509	21117
Number of sequences with length > 80 bp & SSR motif > 4 repeats	91	4183
Sequences length (Mean-Min-Max)	244-127-731	278-80-548
% Loci identified from several sequences	28,5	14,3
Number of sequences with successful Design of primers	7	266
Ratio of perfect SSR / compound SSR	5-2	185-81
Duration	1 month	2 months (Genoscreen)
Cost	> 2500 euros	1750 euros (Genoscreen)

Table 1 : Description of the microsatellite libraries obtained by the two methods.

RESULTS

We obtained 21,117 sequences for the new technology vs 509 recombinant clones (potentially sequences) for the classical one. In total, 4,183 loci containing microsatellites matching the quality criteria implemented in QDD were detected after the 454 technology (Table 1). For the 454 technology, mean sequence length was of 278 bp (244 bp for classical ones) and maximum was of 548 bp (731 bp for classical ones), 14,3 % of the validated loci corresponded to consensus sequences (28,5 % for the classical ones). Primer design was successful in 6,3 % loci (266 out of 4,183), (7,6 % for classical ones). In both technologies, the primers design targeted more microsatellites with perfect motif than with compound ones.

CONCLUSIONS

Even with a stringent selection of loci imposed by the parameters chosen throughout the analysis, the total number of microsatellites isolated by using the new 454 technology is much higher than the number of SSR obtained following a classical method. The use of multiplex enrichment with 8 probes made the SSR enrichment easier and generalist. Moreover, the analysis of the large dataset obtained, using bioinformatic programs like QDD also enabled to sort and discard loci found at multiples sites in the genome.

The data presented here confirmed that the microsatellite isolation through the 454-GS-FLX Titanium pyrosequencing of enriched library can be successfully applied to a non model taxa at much lower cost and more rapidly than by using a classical method.

Acknowledgements

The autors thank Thibaut Malausa (INRA, UMR 1301 IBSV INRA/UNSA/CNRS, Sophia Antipolis, France), André Gilles, Caroline Costedoat, Vincent Dubut, Nicolas Pech & Emese Meglècz (Aix-Marseille University, CNRS, IRD, UMR 6116-IMEP, France), Stephanie Ferreira, Hélène Blanquart, Stephanie Duthoy (Genoscreen, Genomic Platform and R&D, Lille, France) for their valuable collaboration to this project.

References

Kijas J.M., Fowler J.C., Garbett C.A., Thomas M.R., 1994. Enrichment of microsatellite from the citrus genome using biotinylated oligonucleotide sequences bound to strepatvidin coated magnetic particles. *Biotechniques.* 16(4) : 656-60, 662.

Malausa T., and al., 2010. High-throughput microsatellite isolation through 454 GS-FLX titanium pyrosequencing of enriched DNA libraries. *Molecular Ecology Resources* submitted.

Meglécz E., Costedoat C., Dubut V., Gilles A., Malausa T., Pech N., Martin J.F., 2010. QDD : a user-friendly program to select microsatellite markers and design primers from large sequencing projects. *Bioinformatics* . 26,3 : 403-404.